OTS: 60-11,684

JPRS: 2729

9 June 1960

# THE USE OF ELECTRONIC COMPUTERS IN MEDICAL STATISTICS

- USSR -

by A. D. Voskresenskiy and A. I. Prokhorov



Distributed by:

OFFICE OF TECHNICAL SERVICES
U. S. DEPARTMENT OF COMMERCE
WASHINGTON 25, D. C.



U. S. JOINT PUBLICATIONS RESEARCH SERVICE
205 EAST 42nd STREET, SUITE 300
NEW YORK 17, N. Y.

JPRS: 2729

CSO: 3812-N/a

# THE USE OF ELECTRONIC COMPUTERS IN MEDICAL STATISTICS

[This is a translation of an article written by A. D. Voskresenskiy and A. I. Prokhorov in Sovetskoye Zdravookhraneniye (Soviet Health), Vol. XIX, No. 2, Moscow, 1960, pages 25-32.]

The electronic computers produced in recent years constitute the greatest achievement of the human mind, and open a new era in science and technology. The theoretical and practical basis for the use of electronic computers in medicine is their ability to perform mathematical and logical computations of enormous complexity, in the shortest time, operating on more quantitative data than the human memory can store. Any problem presented as a continuous sequential chain of logical operations can be solved by an electronic computer.

From well-known reports describing the volume of work in public health planned for the next seven years, it is apparent how great a need exists for the study and analysis of the quantitative aspects of various phenomena, and for broad and intense statistical investigations encompassing the variety of relationships with which medicine has to deal. Statistical processing of observations is an indispensable condition for planning prevention and hospitalization measures. Establishing hygienic norms for work and rest; determining the value of medicinal remedies, vaccines, serums, and methods of surgical treatment; measuring the effect of industrial and domestic factors on a person's health, organizing drugs supply -- all these are impossible to do without applying statistical methods to the information obtained in the course of an observation. The greater the number of phenomena and relationships subject to statistical analysis, the more certain and accurate are the conclusions which can be made with regard to the causes and regularity of diseases.

The uniformity of principles and quality of medical facilities adopted in Soviet public health yield all the conditions for organizing perfect statistical research systems. However, in the solution of some important problems concerning clinics and disease prevention, statistical study of facts is not given sufficient attention and facilities. Weak knowledge and utilization of statistical laws result, in many cases, in an argument concerning one or many treatment methods, or concerning the factors influencing the oscillations in sickness rates -- an argument which may last for years on end without producing evident proofs of the usefulness of any particular view. Such a situation arises as a result of insufficient quantitative observations, or incomplete study of conditions which may influence the conclusions. Application of exact research methods and realization of the mathematical and logical possibilities of electronic computers may serve as the basis for a radical improvement in the position of statistics in medicine and biology.

The application of electronic computers to the solution of statistical problems, as experienced in the U.S.S.R. and abroad, proves their enormous possibilities in this direction. The effectiveness of computer utilization stems from the automation of the most laborious statistical processes -- classification of incoming materials into groups and production calculations -- as well as the tremendous acceleration of statistical data processing, paralleled by the reduction of the work required for accomplishing similar tasks by ordinary methods.

At the present stage, the development of computational techniques for statistics opens new and broad perspectives, owing to the possibilities of electronic computers. Numerical analyzers of the electromechanical type, which are currently in wide use, have a number of technical peculiarities which limit the range of problems solvable by such machines. They can perform complex combinations of arithmetical operations (summation, calculation of various indicators, etc.), but they cannot serve for the complete automation of a statistical process. Electronic computers possess a basically different property in that they operate on the basis of alogorithms, i.e., precise instructions, which, as will be shown, yield correct solutions to all problems in a definite form. Electronic computers can accomplish any logical task for which an alogorithm can be found and formulated. The field of medical statistics is particularly amenable to alogorithmization, since the framework of statistical research processes is constructed on mathematical methods of analysis and on a rigorous, logical

system of inferences. It is appropriate to present an example illustrating the inferiority of conclusions resulting from the lack of complete statistical control. Consider one of the tables produced by a group of authors in the journal "Voprosy Virusologii", No. 1, 1959, page 38.

Incidence of Influenza in June-October 1957
per 100 persons of corresponding age

| | Age of subjects | | | | Total | As given by town registry |
|---|---|---|---|---|---|---|
| | 0-2 | 3-6 | 7-14 | Adults | | |
| Number of subjects | 127 | 159 | 432 | 2751 | 3469 | |
| Incidence of influenza: | | | | | | |
| Absolute number | 59 | 95 | 244 | 1277 | 1675 | |
| % | 46.5 | 59.6 | 56.4 | 46.4 | 48.3 | 37.0 |

The authors find the susceptibility to influenza to be approximately the same in all age groups; to corroborate this the given table, based on the results of a population questionnaire, is produced. At first sight, the incidence percentages in the various groups seem indeed similar. However, if the mean differential error is computed for the percentage ratio in the first two groups, the conclusions reached by the authors become doubtful. The mean differential error for the rates in the first and second groups equals $\sqrt{19.6 + 15.17} = \pm 5.9\%$, and twice the mean differential error (11.8%) is less than the difference between the indicators (59.6 - 46.5 = 13.1%); consequently, the difference between the indicators in these two groups is, with high probability, not accidental but caused by unequal susceptibility of the groups to the influenza virus, or by differences in some factors or others which affect the incidence. The authors assume equivalent conditions for both groups, in which case it is even less feasible to explain fortuity. If conditions are considered nonequivalent, then it is impossible to compare the incidence, since in this case comparison would be incomplete. A similar result is obtained when the mean differential error is computed between the second and fourth groups.

The example given shows that statistical methods should occupy a more appropriate place in medicine.

The problem of automatic diagnosis of many diseases likewise requires a radical improvement in clinical statistics, since only by the integration of all the reliable information concerning the frequency and significance of symptoms is it possible to establish diagnosis formulas for given complex situations.

It can be assumed that in the future clinical statistics will be distinguished as an independent division of statistics, owing to the importance and specific value of its problems.

The variety and quantity of problems to be solved by statistical medicine indicate the need for strict standardization of statistical material collection, and for automatic processing of this material. Success in this matter, as well as considerable saving in time, material and labor, can be provided by the use of electronic computers. With the aid of computers it is possible to automatize not only a computational process but also the logical aspects of such a process, i.e., choosing the course of calculations in accordance with the results of intermediate operations, as applied to the entire arsenal of mathematical methods, and establishing dependence criteria among the various factors. By these means, answers would be obtained in a form most convenient for subsequent comparison and for planning action in accordance with contemplated rules.

Alogorithmization of statistical problems can be realized only through the joint effort of medical statistics specialists, representing various areas in medicine, and mathematicians-programmers. In the establishment of programs for machines of a specified type, the pariticpation of an electronic engineer is also obligatory. The doctors' responsibility should encompass the following: 1) choosing for the statistical research to be undertaken that having the greatest significance and highest priority in the automatization process; 2) establishing the form of the reports at each stage of their passage through the central statistical organ, while sharply curtailing the reporting operations at other stages; 3) controlling the significance of statistical indicators and correlating theoretical conclusions with actual relationships of observed phenomena; 4) studying the logical structures and sequential process of statistical processing and analyzing the data obtained to produce logical schemes for automatic analysis.

Consider the automatic mechanism for summarizing statistical observations, in the example of processing "patient cards" -- cards filled for discharged hospital patients.

Analogously, it is possible to process, with the aid of computers, registration documents and reports such as sickness questionnaires, death certificates, etc. Every task is composed of three basic steps: 1) collecting observations; 2) summarizing them in tables; 3) selecting necessary indicators, clarifying and exploring interrelationships of various types.

The automatization of the first step -- collecting observations -- is not feasible, at least not for the existing system of registration and reporting.

The second step consists of summarizing statistical observations in tables. As a result of this step we should obtain tables filled with classified and totaled observations, as well as various types of relative (percentage) indicators. The table format, the symbols, and the intervals for the groupings must be established in advance and included in the machine program. This phase of the task is quite difficult, but compensation lies in the fact that a well established system of tables and indicators may serve, without major reconstruction, for a number of years, and has to be merely expanded in new programs intended for computing newly discovered relationships and rules.

At present, the utilization of computers of electromechanical type for this purpose constitutes an enormous task. Each patient card or any other document must be read by a doctor, each symbol must be designated in cipher and then converted into a perforation code. The entire reverse side of the card of a hospitalized patient, including clinical and pathological-anatomical diagnosis of the entire case, additional illnesses, complications, and names of operations, can be ciphered only by an extremely well-informed doctor. It would suffice to say that in Leningrad, the compilation of hospital records comprises 460 diseases, which is by no means the limit*. The diagnosis designated on a card cannot coincide precisely with the formularized record, and it is necessary to establish the form into

*Statistical analysis compiled from hospitalized patients in Leningrad in 1955, released by the Scientific Methods Bureau of Sanitary Statistics of Lengorzdravotdel (Medgiz, 1958).

which it must be brought in order to facilitate ciphering.7 Consequently, it is necessary to read, cipher and convert into a perforation code the data contained in all cards (in the Leningrad example, 300,000 cards).

If we assume that the processing of each card would require 5 minutes, then the processing of only 300,000 cards would consume 25,000 working hours or, considering a single person and a 6½-hour working day, 4,000 working days. It is understandable that the Leningrad Bureau of Statistics, even with mechanical tabulation of the material, could publish its report only after two years of work.

Briefly speaking, the titanic labor required for the preparation of statistical material as machine input, constitutes the basic obstacle to processing the majority of registrations and reports that are of practical interest to medical statistics.

Only the utilization of electronic computers for this purpose would permit the automatization and would considerably accelerate the processing of tens of millions of documents that currently have no particular usefulness, being stacked in archives of many institutions.

Finally, the third step of the statistical investigation -- the analysis of indicators, their comparison in time and space, and the initiation of a deep statistical analysis by detecting changes which exceed the probability of the percent differential due to accidental causes -- all that can be fully automatized thanks to the logical capabilities of computers. Such an analysis (and its automatization) is especially important in large-scale territories, when one can assume a number of factors affecting sickness rates differently in the various individual regions.

The advantage of computers is due to the fact that they facilitate automatic analysis of relationships associated with various combinations of these factors, towards a conclusion concerning a comparative and correlative dependence, if such a dependence exists. To that end, the computer has to be provided with sub-programs of operations and commands, constructed by deciding which factors should be selected to establish correlational relationships.

Returning to the collective work of the Leningrad authors, it should be noted that not one of the tables in this report would require, in the summarizing process, logical manipulations of results obtained from intermediate computations. Each computation is performed in a strictly

routine fashion from the particular to the general, i.e., independently of the age indicators or of the number of bed-days in the individual hospitals from which the over-all age indicators and over-all average number of bed-days per hospital and per age group could be calculated. The results, when compared with those of the preceding year (chronologically), yield the conclusion that some of the indicators have subsided. In cases where indicators are changed for the best, it is possible to conclude that attention should be concentrated on the diagnosis of some pathological area, and so forth. Only in one article are figures produced concerning various indicators for the number of bed-days in some hospitals for a given disease. No attempt has been made to analyze the sharp differences in the treatment dates of one disease or another (in patients of the same kind) in the various hospitals.

As a matter of fact, it is impossible to obtain the causes for the above differences from statistical analysis of patient cards, since treatment dates are substantially affected by many factors not accounted for in the records and in the statistical forms.

However, one could attempt to establish the relationship between treatment dates and age of patients (or generally, characteristics of patients treated in given institutions), dates of patient admissions, period extending from the time of admission to the time of operation, complications, responses, coincidences, diagnostical discrepancies, etc.

On the other hand, special analysis of an individual medical institution is possible (or, rather, possible within the probability sense of the true conclusions) only for those victims of any given disease, the number of which would yield at least some scores of observations in the investigated medical institution (so that the indicator differential would exceed twice the average error differential of the percentage ratio). Such a selection of hospitals and investigation of relationships of various orders can be performed automatically by programmed computers.

It is understood that one may obtain the relationships with controllable accuracy or reliability. The value of statistical research by formal methods (conducted according to the same principles as ordinary methods, but in considerably larger volumes) is evident, since the presentation of the relationships makes the research efforts useful in subsequent processing, in the selection and analysis of observations, which will be performed in a more rational

| Encoded form of patient cards | |
|---|---|
| – | Does he live in given locality? |
| o – – o | Number of hours after start of illness (submission to trauma) |
| – | was special surgical evidence given in station? |
| – o o – | operation performed |
| – – o | confinement time in institution |
| – o – o | age group |
| – | sex symbol |
| o | response symbol |
| – o o – o | Diagnosis of the referring institution |
| – – – o – o | station address |

manner, and in eliminating laborious, unnecessary computations where a formal relationship among various effects is not apparent. Any mass of information introduced into a computer in a qualitative form is utilized maximally, as compared to any other methods, both with respect to quantitative performance and execution time (the latter being especially important in epidemology, where operational analysis is required).

*  *

*

We shall consider the sequential and logical operation of machines* in the example of processing patient cards.

The data included in the card must be encoded (see table) and introduced into the machine. This

*Naturally, subsequent presentation of the material would require some elementary knowledge, on the part of the reader, of the principles of computer operation and construction. Since we cannot discuss these problems here, we refer the reader to our article, published in "Voyenno-meditsinakom Zhurnale," No. 6, 1959.

process can be completely automatized. Computers operate numerically in the so-called binary number system, through which any number in the decimal system can be expressed by a corresponding combination of ones (1) and zeros (zero). Each number is stored in the machine memory system, in a particular location called a "cell". In each cell of the memory system, a number is stored consisting of ones and zeros, and representing all the data in the patient card in a coded form.

After the data of all the cards to be processed (practically, there may be as many of these as desired) are recorded in the machine "memory", processing starts in accordance with a pre-established program. Admittedly, we must establish tables and generate a number of indicators which are of interest (see table of observations summary). In this case, the machine would operate initially in accordance with a logical diagram (see diagram of a address searching). Each number (as given in some patient card) is picked up from the corresponding cell of the memory system (numbers are picked up in order), and operated on by the arithmetic system of the machine. In the arithmetic system, a comparison operation is carried out between this number and pre-established numbers; rather than the entire number, only the necessary portion is compared (first, the diagnosis symbol). In case of coincidence, automatic switching is effected to the next operation -- comparison with the next symbol -- and so on, until the final result is recorded in a machine memory location stipulated by the program. In case of lack of coincidence of any particular symbol, the entire number is restored to its cell in the machine memory, and the next number is picked up. Thus, sequentially, step by step, numbers can be selected from all the remaining cells and the necessary table established. Subsequently (after switching to another program), the computation of the necessary indicators and determination of the corresponding relationships can be initiated.

Performing these operations manually would prove too slow and cumbersome. For machines, however, these operations are most suitable. Regardless of how efficient people may be, even with the aid of desk calculators, their operation speeds would always be considerably lower than that of a computer, since each such operation takes, in a computer, only a millionth of a second. Also, the size of a practical program may be as large as desired. Switching from one of the programs executed by the machine to another may be realized either automatically by a special program established for this purpose (in this case the special program is called a "sub-program"), or

Table of observations summary as grouped in computer

| Diagnosis (basic) | Outcome of disease | Sex | Age (years) | Number of patients | Number of bed-days | Other indicators |
|---|---|---|---|---|---|---|
| Measles | Discharged | Male | 0-1 | 14 | 380 | --- |
| | | | 2-2 | 7 | 180 | --- |
| | | | --- | --- | --- | --- |
| | | | --- | --- | --- | --- |
| | | | 35-40 | --- | --- | --- |
| | | Female | 0-1 | --- | --- | --- |
| | | | 1-2 | --- | --- | --- |
| | | | --- | --- | --- | --- |
| | | | --- | --- | --- | --- |
| | | | 35-40 | --- | --- | --- |
| | Died | Male | 0-1 | --- | --- | --- |
| | | | 1-2 | --- | --- | --- |
| | | | --- | --- | --- | --- |
| | | | --- | --- | --- | --- |
| | | | 35-40 | --- | --- | --- |
| | | Female | 0-1 | --- | --- | --- |
| | | | 1-2 | --- | --- | --- |
| | | | --- | --- | --- | --- |
| | | | --- | --- | --- | --- |
| | | | 35-40 | --- | --- | --- |

conditionally, depending upon computation results obtained.]

The diagram presented does not by any means exhaust the logical capabilities of computers; although it is the result of serious work, it cannot be claimed as complete and indisputable, and is given here only as a visual illustration of the advantages and possibilities of computers in processing documents of considerable interest to medical statistics. Thanks to electronic computers, statistics should and can extend its range of effective application not only in width but also in depth, penetrating with its methods into all fields of theoretical and practical medicine. The results of statistical research should occupy a proper position not only in specialized statistical literature but also in clinical, physiological, epidemological and other publications. Clinical journals should publish statistics research papers, with examples which reveal the inadequacy of individual works, the authors of which construct their conclusions on the basis of incomplete statistical data; they should summarize and analyze, by statistical methods, the results of observations based on aggregates of facts. Medical and biological statistics must be guided, in their conclusions, by the joint knowledge of doctors, biologists and mathematicians.

The application of computers in statistical research permits us to perform computations in an extremely short time; thus statistics becomes an effective weapon in the operational struggle for the reduction of disease and the improvement of human health. The various interacting causes and consequences, the abundance and variability of factors which influence each other, the probabilistic character of many relationships, the necessity for processing enormous quantitative data -- all these make the problems of statistical research of biological behavior extremely difficult and complex. Precisely for this reason, the intricate methods required for solving such problems must be applied through the use of all available electronic achievements, and above all electronic computers.

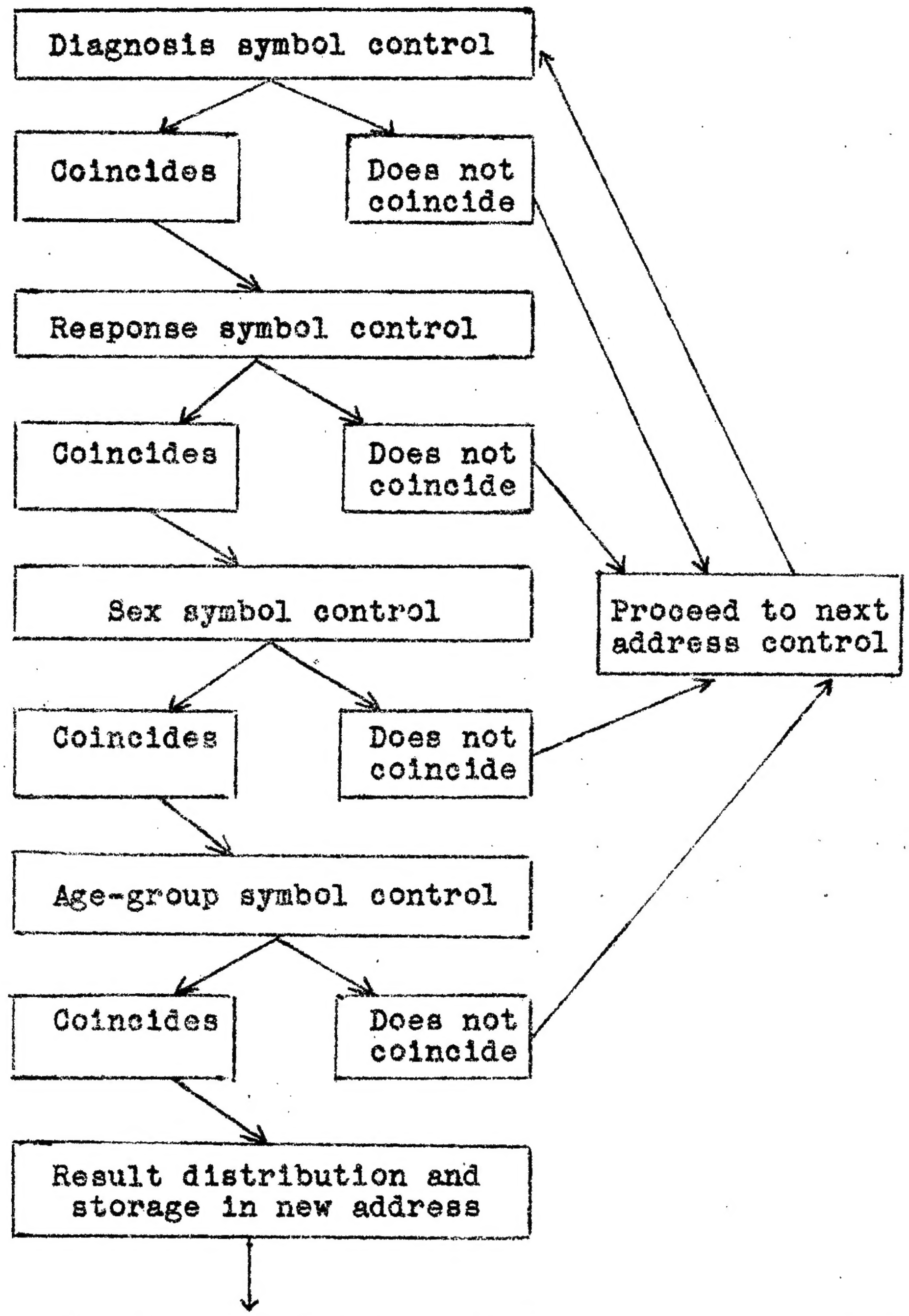


Diagram of address-searching in machine memory

Submitted 8 October 1959.

- END -

FOR REASONS OF SPEED AND ECONOMY
THIS REPORT HAS BEEN REPRODUCED
ELECTRONICALLY DIRECTLY FROM OUR
CONTRACTOR'S TYPESCRIPT